

3D-QSAR studies of 8-substituted chromen-4-one-2-carboxylic acid derivatives as potent agonists for the orphan G protein-coupled receptor 35

Taigang Liang · Chaoqun Yan · Li Yang · Mengnan Hu · Shurong Ban · Qingshan Li

Received: 1 July 2014 / Accepted: 24 October 2014 / Published online: 6 November 2014
© Springer Science+Business Media New York 2014

Abstract GPR35, an orphan G protein-coupled receptor, has attracted much attention as a novel therapeutic target for the treatment of diabetes, hypertension, etc. Recently, 8-substituted chromen-4-one-2-carboxylic acid derivatives were identified as potent and selective agonists for human GPR35. In the present study, the three-dimensional quantitative structure–activity relationship (3D-QSAR) models were developed for a series of 8-substituted chromen-4-one-2-carboxylic acid derivatives using comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA), and Topomer CoMFA techniques implemented in the SYBYL software packages. The statistically significant models were obtained with 30 compounds in training set by ligand-based atom-by-atom matching alignment, which were further validated by a test set of eight compounds. The CoMFA model resulted in cross-validated coefficient (q^2) value of 0.610 using 4 components, non-cross-validated coefficient (r^2) value of 0.918 with estimated F value of 69.917, and standard error of estimate (SEE) of 0.352. While the CoMSIA model combined with steric,

electrostatic and hydrophobic fields were finally selected ($q^2 = 0.646$, $r^2 = 0.800$, $F = 53.852$, $SEE = 0.489$, $N = 2$). For the Topomer CoMFA model, the better statistics were obtained based on fragment units ($q^2 = 0.746$, $r^2 = 0.979$, $F = 146.294$, $SEE = 0.175$, $N = 7$). Furthermore, the contour maps obtained from 3D-QSAR studies were appraised for activity trends for the compounds analyzed. The results indicate that steric, electrostatic, and hydrophobic substituents play a significant role in the agonist activity. The data generated from the present study will further help design novel, potent, and selective agonists for GPR35.

Keywords 8-Substituted chromen-4-one-2-carboxylic acid derivatives · GPR35 agonists · 3D-QSAR · CoMFA · CoMSIA · Topomer CoMFA

Abbreviations

3D-QSAR	Three-dimensional quantitative structure–activity relationship methods
CoMFA	Comparative molecular field analysis
CoMSIA	Comparative molecular similarity indices analysis
GPCRs	G protein-coupled receptors
GPR35	G protein-coupled receptor 35
LOO	Leave-one-out procedure
L5O	Leave-five-out procedure
N	Optimal number of components
PLS	Partial least square method
PRESS	Predictive sum of squares
q^2	The cross-validated correlation coefficient
q_{L5O}^2	Cross-validated correlation coefficient for leave-five-out
r^2	Non-cross-validated correlation coefficient
SEE	Standard error of estimate

Taigang Liang and Chaoqun Yan have contributed equally to this work.

T. Liang · C. Yan · M. Hu · S. Ban · Q. Li (✉)
School of Pharmaceutical Science, Shanxi Medical University,
No 56, Xinjian Nan Road, Taiyuan 030001, Shanxi,
People's Republic of China
e-mail: sxlqs2012@163.com

T. Liang
e-mail: ltaigang@gmail.com

L. Yang
Shanxi University of Traditional Chinese Medicine, No 89, Jinci
Road, Taiyuan 030024, Shanxi, People's Republic of China

Introduction

The G protein-coupled receptors (GPCRs) constitute a large family whose members are involved in numerous physiological functions and represent more than 30 % of all pharmaceutical drug targets (Pal *et al.*, 2012). While the sequencing of human genome has led to the discovery of novel GPCRs, many of them are orphan receptors for which the natural ligands remain undefined (Milligan, 2002). As one of the orphan GPCRs, GPR35 was initially described in 1998 sharing homology with GPR55, the lysophosphatidic acid receptors LPAR₄, LPAR₅, and LPAR₆, and the nicotinic acid receptor HM74 (Mackenzie *et al.*, 2011). GPR35 has been found to be expressed in various tissues, especially in gastrointestinal, immune, and nervous systems (Cosi *et al.*, 2011; Ohshiro *et al.*, 2008; Taniguchi *et al.*, 2006). Meanwhile, accumulated evidence has confirmed that the activation of GPR35 plays a key role in some physiological and pathological progresses (Mackenzie *et al.*, 2011). For example, GPR35 agonists can regulate insulin release (Jenkins *et al.*, 2011), lower blood pressure and modulate immune responses (Min *et al.*, 2010; Wang *et al.*, 2006). Thus, GPR35 has been recognized as a novel therapeutic target in conditions that range from diabetes and hypertension to asthma. In the process of searching for GPR35 agonists, kynurenic acid, a tryptophan metabolite, was firstly described to be a possible endogenous ligand (Wang *et al.*, 2006). However, because of its low potency, more and more attentions have been focused on synthetic agonists. In the past decade, several synthetic GPR35 agonists have been reported in the literature such as the antiallergic drug cromolyn (Taniguchi *et al.*, 2006), zaprinast (Yang *et al.*, 2010), thiazolidinylidene-methylbenzoic acid derivatives (Neetoo-Isseljee *et al.*, 2013), and so on. Unfortunately, most of these existing GPR35 agonists still have some major drawbacks including low potency and limited selectivity. It is noteworthy that most recently a series of 8-substituted chromen-4-one-2-carboxylic acid derivatives, which shared a substructure with cromolyn, have been identified as potent and selective GPR35 agonists (Funke *et al.*, 2013; Thimm *et al.*, 2013).

Nowadays, various computational tools have been increasingly employed in rational drug discovery process to design new molecules and to optimize structure (Heikamp and Bajorath, 2013). Commonly used computer modeling techniques mainly include structure-based drug design and ligand-based drug design. Structure-based drug design relies on knowledge of the three-dimensional structure of the biological target obtained through methods such as x-ray crystallography or NMR spectroscopy. In the absence of detailed structural information about the receptor, ligand-based approaches such as three-dimensional quantitative structure–activity relationship (3D-QSAR) methods could be more

effective (Wilson and Lill, 2011). 3D-QSAR methods can facilitate the correlation of the three-dimensional structures of potential drug molecules with their biological activity and in turn help to predict the activity of new molecules prior to synthesis (Chandrasekaran *et al.*, 2004). Until now, the three-dimensional structure of GPR35 is not available. In the present paper, a series of 8-substituted chromen-4-one-2-carboxylic acid derivatives were used to develop 3D-QSAR models using comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA), and Topomer CoMFA methods. The obtained models can be used to identify the structural features essential for enhancing their activities and subsequently can enable the design of the new more potent GPR35 agonists.

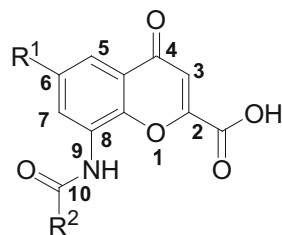
Materials and methods

Dataset

In the present work, thirty-eight 8-substituted chromen-4-one-2-carboxylic acid derivatives were taken from the published work of Funke *et al.* (2013). The human GPR35 agonist activities were tested by a β -arrestin recruitment assay, the most commonly used assay system to pharmacologically characterize GPR35 ligands. In this assay, stock solutions of the compounds were prepared in DMSO and further diluted in cell plating medium or phosphate buffered saline. So compounds could be regarded to exert their activities in neutral forms. Subsequently, the EC₅₀ values were converted into the corresponding pEC₅₀ values by the formula $pEC_{50} = -\log EC_{50}$. The pEC₅₀ values have a span of three log units providing a broad and homogenous dataset for 3D-QSAR study. The structures of the compounds and their biological data are given in Table 1. The test set of eight compounds is chosen at random from the original dataset, which is used to determine the external predictivity of the resulting 3D-QSAR models. The remaining 30 compounds are treated as a training set and used to derive the 3D-QSAR models (Table 1).

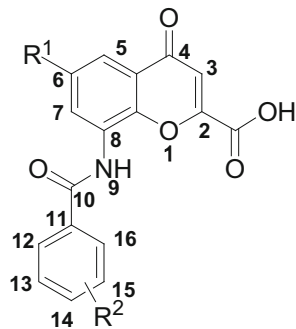
Molecular modeling and alignment

The molecular modeling and calculations were performed using SYBYL program (SYBYL- X2.0, Tripos Inc., St. Louis, MO, USA) on windows operating system. All the molecular structures were built in SYBYL and then the energy minimization was performed using Tripos force field with a distance-dependent dielectric function and Powell conjugate gradient algorithm with a convergence criterion of 0.01 kcal/mol Å using 1,000 iterations. Partial atomic charges were calculated using the Gasteiger-Huckel method (Clark *et al.* 1989).

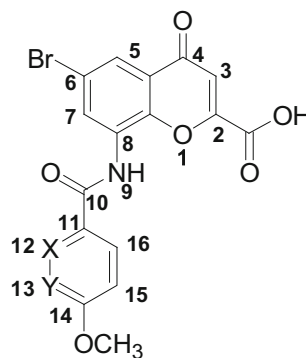
Table 1 Structures and biological activities of the training and test sets of compounds

1-8

Compound	R ¹	R ²	EC ₅₀ (μM)	pEC ₅₀
1	H	Methyl	24.1	4.618
2 ^a	H	Ethyl	13.4	4.873
3	H	2-Methoxy-2-oxoethyl	20.2	4.695
4	H	Cyclohexyl	4.20	5.377
5	H	Phenyl	4.91	5.309
6	H	Benzyl	8.96	5.048
7 ^a	H	2-Naphthyl	4.67	5.331
8	H	2-Quinoliny	1.34	5.873



9-36



37 (X=N, Y=CH)

38 (X=CH, Y=N)

	R ¹	R ²			EC ₅₀ (μM)	pEC ₅₀
		<i>ortho</i>	<i>meta</i>	<i>para</i>		
9	H	H	CH ₃	H	5.44	5.264
10	H	H	NO ₂	H	6.79	5.157
11 ^a	H	H	H	CH ₃	1.38	5.860
12	H	H	H	CF ₃	15.5	4.810
13	H	H	H	OCH ₃	0.346	6.461
14	H	H	H	Br	0.804	6.095
15 ^a	F	Cl	Cl	H	0.382	6.418
16	F	Cl	H	Cl	0.017	7.770
17	F	H	Cl	Cl	0.117	6.932
18	F	H	H	OCH ₃	0.112	6.951
19 ^a	Cl	H	H	H	0.430	6.367

Table 1 continued

	R ¹	R ²			EC ₅₀ (μM)	pEC ₅₀
		<i>ortho</i>	<i>meta</i>	<i>para</i>		
20	Cl	H	CH ₃	H	1.22	5.914
21	Cl	H	H	NO ₂	1.71	5.767
22 ^a	Cl	H	H	OCH ₃	0.0168	7.775
23	Br	H	H	H	0.303	6.519
24	Br	H	CH ₃	H	0.842	6.075
25	Br	H	H	Cl	0.0251	7.600
26 ^a	Br	Cl	H	Cl	0.0164	7.785
27	Br	H	Cl	Cl	0.0154	7.812
28	Br	H	H	CN	0.399	6.399
29	Br	H	H	OCH ₃	0.0121	7.917
30	Br	H	H	Ethoxy	0.634	6.198
31 ^a	Br	OCH ₃	H	H	1.66	5.780
32	Br	H	OCH ₃	OCH ₃	0.425	6.372
33	Br	H	OCH ₂ O		0.0599	7.223
34	Br	Cl	H	OCH ₃	0.0111	7.955
35	H	H	H	Cyclopropylmethoxy	10.0	5.000
36	H	H	H	Propoxy	4.0	5.398
37					0.144	6.842
38					0.0305	7.516

^a Test set

The alignment of the compound structures is a critical step in 3D-QSAR methodologies, such as CoMFA/CoMSIA. The accuracy of the prediction of CoMFA and CoMSIA models and the reliability of the contour models depend strongly on the structural alignment of the compounds (Cho and Tropsha, 1995). In the present study, the most potent compound **34** was employed as a template, and the rest of the compounds in the training set were aligned to it using the common substructure. Figure 1 describes the common substructure for the alignment which is marked in bold red and the aligned compounds are displayed in Fig. 2.

CoMFA studies

All the aligned compounds were placed in a 3D cubic lattice with a grid spacing of 2 Å which was generated automatically by the sybyl program. CoMFA descriptors were calculated using Tripos force field taking a sp³ carbon probe atom with a Van der Waals radius of 1.52 Å and a positive charge +1 to generate steric (Lennard-Jones 6-12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric at each lattice point. Values of the steric and electrostatic fields were truncated at 30.0 kcal/mol.

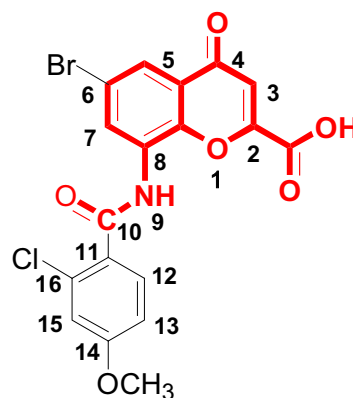


Fig. 1 Chemical structure of compound **34** used as template compound in 3D-QSAR modeling. The common substructure used for compound alignments is represented in **bold red** (Color figure online)

CoMSIA studies

CoMSIA similarity index descriptors were derived according to Klebe et al. (1994) using the same lattice box as that used in CoMFA calculations. Five different similarity fields including steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor

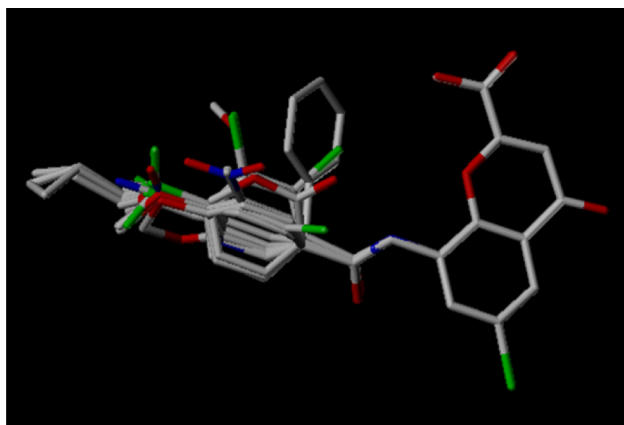


Fig. 2 Alignment of 30 compounds of training set for 3D-QSAR studies

interactions were calculated using the sp^3 atom with charge +1 and radius of 1 Å. CoMSIA similarity indices ($A_{F,K}$) for the molecule j and the atom i at grid point q were calculated as follows:

$$A_{F,K}^q(j) = - \sum_{i=1}^n \omega_{\text{probe},k} \omega_{i,k} e^{-ar_{iq}^2}$$

where $\omega_{i,k}$ is the actual value of the physicochemical property k of atom i ; $\omega_{\text{probe},k}$ indicates the probe atom with charge +1, radius 1 Å, hydrophobicity +1, hydrogen bond donor and acceptor property +1; a is the attenuation factor; r_{iq} is the mutual distance between the probe atom at grid point q and atom i of the test compound. In this paper, steric indices are related to the third power of the atomic radii, electrostatic descriptors are derived from partial atomic charges, hydrophobic descriptors are derived from atom-based parameters, and H-bond donor and acceptor indices are obtained by a rule-based method based on experimental results. The default value of 0.3 was used as the attenuation factor (a) for the Gaussian-type distance r_{iq} .

Topomer CoMFA studies

The Topomer CoMFA descriptors were calculated using sybyl-X2.0 software package. Topomer CoMFA is an alignment-independent 3D-QSAR that combines the topomer search technology with the conventional CoMFA method (Cramer, 2012). The Topomer CoMFA model can be generated by splitting the compounds into fragments, topomerically aligning each fragment, and calculating steric and electrostatic field descriptor values for the topomerically aligned fragments to create a CoMFA table with the field descriptor values. Identifying the R-groups for the training set compounds was the important step for

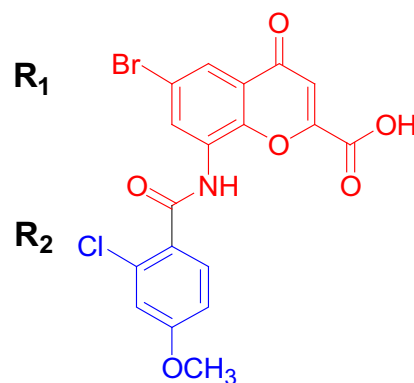


Fig. 3 Fragmentation pattern (R_1 and R_2) for all compounds of dataset in Topomer CoMFA analysis. R_1 fragment is represented by the red color and R_2 fragment is denoted by the blue color (Color figure online)

Topomer CoMFA. In this study, all compounds of dataset were splitted into two fragments by cutting a single bond, shown as R_1 (red) and R_2 (blue) groups in Fig. 3.

Partial least square (PLS) analysis

Partial least squares (PLS) approach, an extension of the multiple regression analysis, was used to derive the 3D-QSAR models. Column filtering was set to 2.0 kcal/mol to improve the analysis and reduce the noise. In PLS regression analysis, the CoMFA, CoMSIA, and Topomer CoMFA descriptors were used as independent variables and biological activity (pEC_{50}) as dependent variables.

To check statistical significance of the models, cross-validation analysis performed by the leave-one-out (LOO) procedure was carried out to identify optimum number of components (N), subsequently used to generate the final QSAR models. The optimal numbers of components were selected on the basis of the highest cross-validated correlation coefficient (q^2), which is defined as follows:

$$q^2 = 1 - \frac{\sum (Y_{\text{prep}} - Y_{\text{exp}})^2}{\sum (Y_{\text{exp}} - Y_{\text{mean}})^2},$$

where Y_{pred} , Y_{exp} , and Y_{mean} are the values (pEC_{50}) for the predicted activity, experimental activity, and mean activity, respectively. $\sum (Y_{\text{prep}} - Y_{\text{exp}})^2$ is the predictive sum of squares (PRESS). All the results were graphically represented by field contour maps, where the coefficients were generated using the StDev*Coeff field type.

To further assess the robustness and statistical confidence of the derived 3D-models, the leave-five-out (L5O) procedure was also utilized by deleting five compounds from the training set and the corresponding $q^2(q_{L5O}^2)$ was reported.

Table 2 PLS statistical results of CoMFA, CoMSIA, and Topomer CoMFA models

Parameters	CoMFA	CoMSIA	Topomer CoMFA
q^{2a}	0.610	0.646	0.746
r^{2b}	0.918	0.800	0.979
N^c	4	2	7
F^d	69.917	53.852	146.294
SEE ^e	0.325	0.489	0.175
r_{pred}^{2f}	0.892	0.473	0.722
q_{LSO}^{2g}	0.574	0.535-	–

^a Cross-validated correlation coefficient

^b Non-cross-validated correlation coefficient

^c Optimal number of components

^d F test value

^e Standard error of estimate

^f The predictive correlation coefficient

^g Cross-validated correlation coefficient for leave-five-out

Predictive ability of the 3D-QSAR models

The predictive ability of 3D-QSAR model was determined from test set of eight compounds not included in the model generation. The predictive correlation coefficient (r_{pred}^2) based on the test set compounds is defined as

$$r_{\text{pred}}^2 = \frac{(\text{SD} - \text{PRESS})}{\text{SD}},$$

where SD is the sum of the squared deviations between the biological activity of compounds in the test set and the mean biological activity of the training set compounds, and PRESS is the sum of the squared deviations between predicted and actual activity values for every compound in the test set.

Results and discussion

CoMFA statistical results

The statistical results obtained from standard CoMFA models constructed with steric and electrostatic fields are summarized in Table 2. The LOO cross-validated q^2 is 0.610 with an optimum number of components is 4. The non-cross-validated PLS analysis produced a conventional r^2 of 0.918 with estimated F value of 69.917 and standard error (SEE) of 0.325. These statistical indexes indicated that the CoMFA model has a good internal predictivity.

CoMSIA statistical results

CoMSIA models can be constructed by five different fields including steric, electrostatic, hydrophobic, hydrogen bond

Table 3 The results of CoMSIA models based on different field combinations (final CoMSIA model in bold)

Descriptors	q^2	r^2	N	F	SEE
E	0.471	0.649	2	24.994	0.646
H	0.524	0.827	5	22.957	0.481
S+E	0.561	0.793	3	33.272	0.506
S+H	0.519	0.882	6	28.558	0.407
E+H	0.649	0.788	2	50.307	0.502
E+D	0.467	0.649	2	24.975	0.647
E+A	0.457	0.665	2	26.766	0.632
H+D	0.508	0.830	5	23.387	0.478
H+A	0.471	0.635	2	23.489	0.659
S+E+H	0.646	0.800	2	53.852	0.489
S+E+D	0.557	0.794	3	33.392	0.505
S+E+A	0.533	0.847	4	34.576	0.444
S+H+D	0.505	0.883	6	28.810	0.405
S+H+A	0.495	0.932	8	35.866	0.323
E+H+D	0.646	0.791	2	51.239	0.499
E+H+A	0.646	0.783	2	48.844	0.508
H+D+A	0.508	0.852	5	27.557	0.446
S+E+H+D	0.643	0.802	2	54.845	0.485
S+E+H+A	0.640	0.793	2	51.700	0.497
S+E+D+A	0.530	0.848	4	35.000	0.442
S+H+D+A	0.499	0.899	6	33.955	0.377
E+H+D+A	0.642	0.787	2	49.771	0.504
S+E+H+D+A	0.637	0.796	2	52.665	0.493

donor, and hydrogen bond acceptor. After testing the every single field, CoMSIA models with hydrogen bond donor, hydrogen bond acceptor, and electrostatic fields were statistically poor ($q^2 < 0.1$, data not shown), whereas both electrostatic and hydrophobic fields play an important role. Furthermore, to generate a reliable model, all possible combinations of fields were examined (Table 3). Among these models, the combination of electrostatic and hydrophobic fields derived the highest q^2 (0.649) using 2 components, with an r^2 value of 0.788, a SEE value of 0.502, and an F value of 50.307. However, considering steric field as a key factor, model including electrostatic, hydrophobic, and steric fields was finally selected, which gave similar q^2 (0.646) with 2 components, relative higher r^2 of 0.800 and F (53.852), as well as small SEE (0.489) (bold in the Table 3), suggesting that the CoMSIA model is also a reliable predictor.

Topomer CoMFA statistical results

The topomer CoMFA technique provides a means for an alignment-independent 3D-QSAR approach which is advantageous in not just being alignment-independent, but also in providing the means for automated search for

Table 4 Experimental activities, predicted activities, and residual values of 38 compounds were shown in CoMFA, CoMSIA, and Topomer CoMFA models

Training set	Actual pEC ₅₀	CoMFA		CoMSIA		Topomer CoMFA	
		Predicted pEC ₅₀	Residues	Predicted pEC ₅₀	Residues	Predicted pEC ₅₀	Residues
1	4.618	4.671	−0.053	5.071	−0.453	4.617	0.001
3	4.695	5.063	−0.368	4.552	0.143	4.68	0.015
4	5.377	5.505	−0.128	5.333	0.044	5.462	−0.085
5	5.309	5.486	−0.177	5.096	0.213	5.343	−0.034
6	5.048	5.084	−0.036	5.024	0.024	5.06	−0.012
7	5.331	5.490	−0.159	5.289	0.042	5.229	0.102
9	5.264	4.870	0.394	5.063	0.201	5.136	0.128
10	5.157	4.944	0.213	5.016	0.141	5.067	0.090
12	4.810	4.771	0.039	5.173	−0.363	4.779	0.031
13	6.461	6.527	−0.066	5.903	0.558	6.546	−0.085
14	6.095	5.601	0.494	5.551	0.544	6.33	−0.235
16	7.770	7.460	0.310	7.011	0.759	7.574	0.196
17	6.932	7.279	−0.347	7.039	−0.107	7.094	−0.162
18	6.951	6.996	−0.045	7.188	−0.237	7.075	−0.124
20	5.914	6.068	−0.154	6.188	−0.274	5.952	−0.038
21	5.767	5.951	−0.184	5.790	−0.023	6.009	−0.242
23	6.519	6.761	−0.242	6.392	0.127	6.463	0.056
24	6.075	6.100	−0.025	6.362	−0.287	6.164	−0.089
25	7.600	6.939	0.661	6.854	0.746	7.407	0.193
27	7.812	7.168	0.644	7.331	0.481	7.593	0.219
28	6.399	6.746	−0.347	6.952	−0.553	6.449	−0.050
29	7.917	7.784	0.133	7.212	0.705	7.575	0.342
30	6.198	6.921	−0.723	7.133	−0.935	6.361	−0.163
32	6.372	6.485	−0.113	7.261	−0.889	6.492	−0.120
33	7.223	7.079	0.144	7.222	0.001	7.271	−0.048
34	7.955	7.937	0.018	7.268	0.687	8.236	−0.281
35	5.000	5.045	−0.045	5.646	−0.646	4.902	0.098
36	5.398	5.185	0.213	5.686	−0.288	5.263	0.135
37	6.842	6.713	0.129	7.383	−0.541	6.627	0.215
38	7.516	7.688	−0.172	7.332	0.184	7.474	0.042
Test set							
2	4.873	5.111	−0.238	5.28	−0.407	4.771	0.102
8	5.873	5.842	0.031	5.354	0.519	5.128	0.745
11	5.86	5.736	0.124	5.158	0.702	5.807	0.053
15	6.418	6.777	−0.359	6.28	0.138	7.087	−0.669
19	6.367	6.801	−0.434	6.525	−0.158	6.251	0.116
22	7.775	7.931	−0.156	6.912	0.863	7.363	0.412
26	7.785	7.433	0.352	7.236	0.549	8.073	−0.288
31	5.780	6.211	−0.431	6.639	−0.859	6.947	−1.167

Residues = Actual pEC₅₀ − Predicted pEC₅₀

activity in fragment libraries. In this study, the Topomer CoMFA model was generated based upon split-in-two strategy (Fig. 3), which gave better statistics ($q^2 = 0.746$, $r^2 = 0.979$, $F = 146.294$, $SEE = 0.175$, $N = 7$) compared with CoMFA and CoMSIA models.

Validation of 3D-QSAR models

Generally, cross-validation is used to assess internal predictive power of models. In this study, the indices q^2 and q_{L50}^2 (>0.5) have accounted for internal robustness of these

models. However, models with high values of cross-validated coefficient are not always predictive. For this reason, it is necessary to have an external set for validation. In this work, the external test set of 8 compounds was used to evaluate the predictive ability of the constructed CoMFA, CoMSIA, and Topomer CoMFA models. As shown in Table 2, the r^2_{pred} values of test set for CoMFA, CoMSIA, and Topomer CoMFA models were 0.892, 0.473, and 0.722, respectively. Obviously, CoMFA and Topomer CoMFA may be capable of obtaining more predictable models, compared with CoMSIA studies. The observed activities, the predicted activities using the constructed models, and the residual values for the training set and the test set are listed in Table 4. In all three 3D-QSAR models, the predicted values fell close to the observed values, deviating by not more than 1.0 logarithmic unit except for one of the test set compounds (**31**) in case of Topomer CoMFA. Compound **31** is an outlier (residual more than 1.0) for Topomer CoMFA, whereas in CoMSIA, the residual value is also quite high (0.859). There are several reasons for the presence of outliers including incorrect biological measurement, incorrect parameter value, metabolic inactivation/activation, etc. (Wang *et al.*, 2014). In the present study, since the assay was carried out under the same conditions by Funke *et al.*, this outlier might be due to experimental error. Figure 4 shows the plots of experimental versus predicted activities for both training set and test set of the three constructed models. In summary, the CoMFA and Topomer CoMFA in comparison with CoMSIA can lead to more robust and predictable model.

CoMFA contour maps

In the CoMFA study, the steric fields and electrostatic fields nearly gave the same contribution, accounting for 52.8 and 47.2 %, respectively, which suggests that both fields are critical in explaining the variations of these compounds for GPR35 agonists.

Figure 5a shows the steric contour map for the CoMFA model with the most active compound **34** as a reference. The green contours indicate areas where sterically bulky substituents increase the potency, whereas the yellow contours indicate areas where sterically bulky groups are detrimental to the activity. There is one green contour around 6-position, which can explain well that most of the compounds with bromine in this position have better activity than those of containing only hydrogen atom. Similarly, increasing the size of the substituent from fluorine (compound **18**, $\text{pEC}_{50} = 6.951$) to chlorine (compound **22**, $\text{pEC}_{50} = 7.775$) to bromine (compound **29**, $\text{pEC}_{50} = 7.917$) at 6-position enhances the activity gradually. There is another small green region located near the 14-position of compound **34**. This is a possible reason why compounds **1**, **2**, and **3** lacking bulky group at this

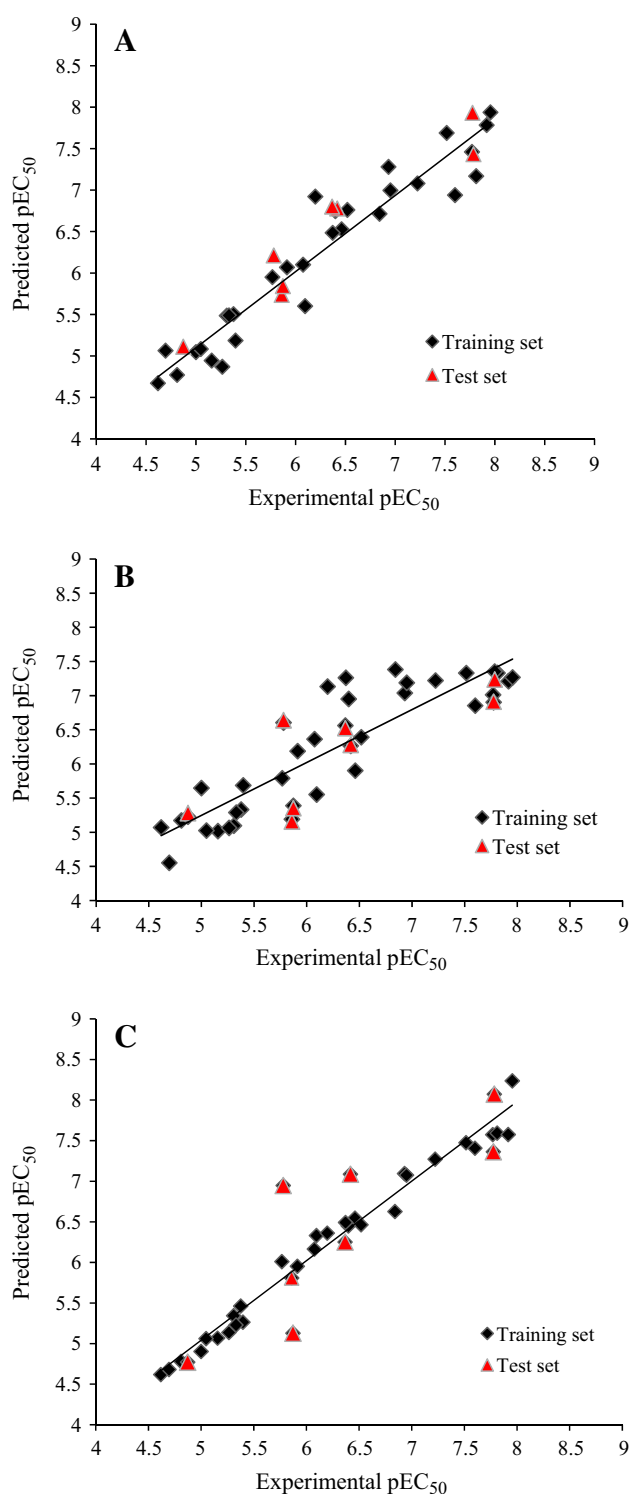


Fig. 4 Plots of the experimental versus predicted pEC_{50} for CoMFA (a), CoMSIA (b), and Topomer CoMFA (c) analysis

position have poor activities (pEC_{50} values are 4.671, 4.873, and 4.695, respectively). R^2 of compounds **1**, **2**, and **3** are small alkyl groups, and introduction of bulkier cyclohexyl substituent can improve potency (compound **4**, $\text{pEC}_{50} = 5.377$). In contrast, all the other compounds have an

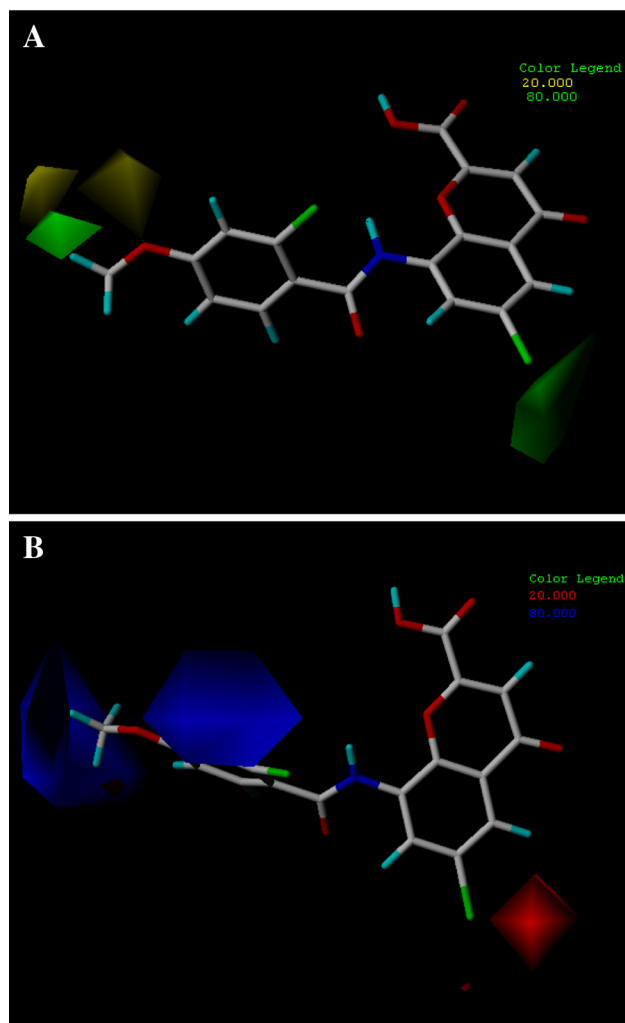


Fig. 5 CoMFA stdev*coeff contour plots for steric (a) and electrostatic (b) fields. Compound **34** was displayed as reference. Sterically favored/disfavored areas are shown in green/yellow, while the blue/red polyhedra depict the favorable site for positively/negatively charged groups. Favored and disfavored levels of these displayed interaction fields were fixed at 80 and 20 %, respectively (Color figure online)

aromatic group at this position, and the majority of these aromatic-substituted compounds displayed the higher efficacies except for compound **12** ($pEC_{50} < 5.0$), which means the existence of an aromatic group at the 10-position is important for activity. Moreover, two sterically unfavored yellow contours were found surrounding the outer sphere of 14-position where too large substituents like cyclopropylmethoxy and propoxy groups (compounds **35** and **36**) decreased the activity. These observations are in agreement with previous SAR studies, and the rank order of potency among compounds with varying substituents in 14-position was as follows: methoxy (**13**) > bromo (**14**) > methyl (**11**) > hydrogen (**5**) > trifluoromethyl (**12**) > propoxy (**36**) \approx cyclopropylmethoxy (**35**) (Funke *et al.*, 2013).

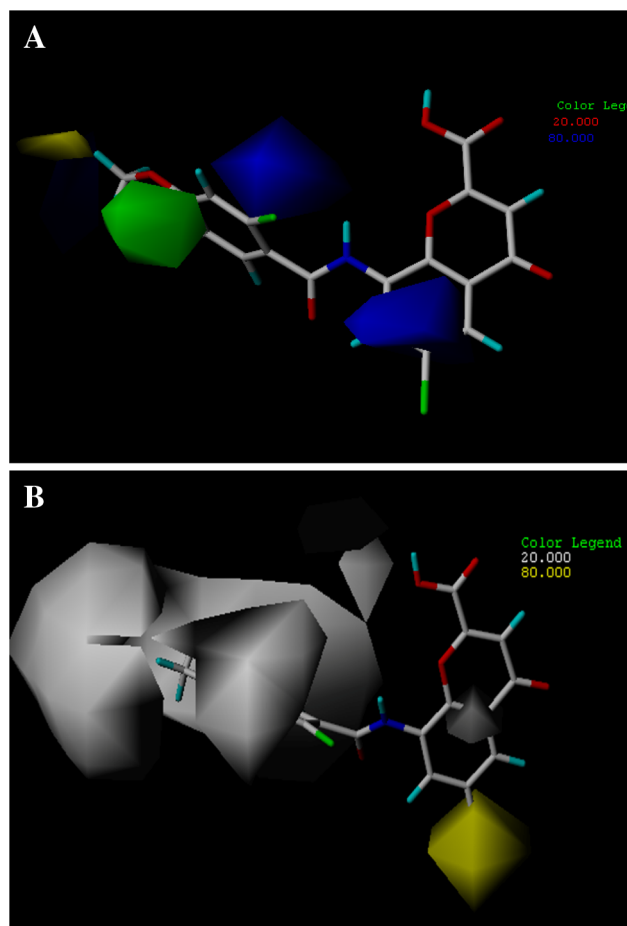


Fig. 6 CoMSIA stdev*coeff contour plots for steric along with electrostatic fields (a) and hydrophobic fields (b). Compound **34** was embedded into the contour maps as reference. Sterically favored/disfavored areas are shown in green/yellow, while the blue/red polyhedra depict favorable sites for positively/negatively charged groups. Yellow/white contours indicate the regions where hydrophobic/hydrophilic groups would enhance the activity. Favored and disfavored levels of these displayed interaction fields were fixed at 80 and 20 %, respectively (Color figure online)

The CoMFA electrostatic contour map is shown in Fig. 5b with compound **34**. A favorable negative red contour region is found around the 6-position. This region indicates that electronegative groups could have a positive influence on the activity. It can explain well that this position is occupied by the electron-rich halogen substituent in most of the highly active compounds. In addition, two large blue regions near the 13- and 14-position suggest that electronegative substituent would decrease the activity. This is in agreement with the fact that the activities of compounds **10**, **12**, **21**, and **28** bearing electronegative substituents such as $-\text{NO}_2$, $-\text{CF}_3$, and CN are relatively lower. Conversely, electron density was decreased due to “ p - π ” conjugation, then compounds **34**

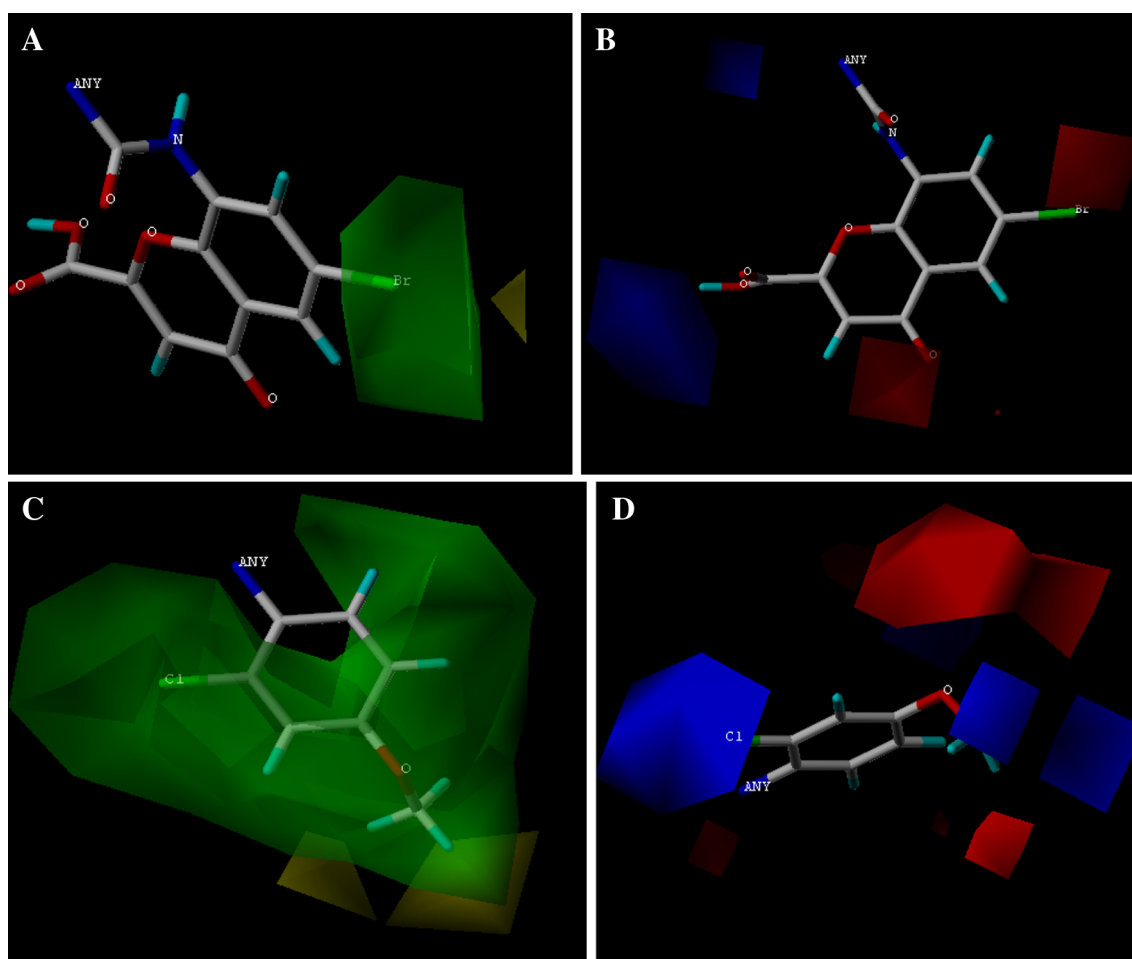


Fig. 7 Topomer CoMFA stdev*coeff contour plots for steric and electrostatic fields. Compound **34** was displayed as reference. Steric contour map for the R₁ fragment (**a**); Electrostatic contour map for the R₁ fragment (**b**); Steric contour map for the R₂ fragment (**c**);

Electrostatics contour map for R₂ fragment (**d**). Sterically favored/unfavored areas are shown in *green/yellow* contour, while the *blue/red* polyhedra depict the favorable sites for the positively/negatively charged groups (Color figure online)

and **29** with methoxy group at *para* position of benzoyl ring display potent activity.

CoMSIA contour maps

The CoMSIA three contributors, namely steric, electrostatic, and hydrophobic fields, based on the PLS analyses are presented as 3D contour plots in Fig. 6. The CoMSIA steric, electrostatic, and hydrophobic fields explain variance of 6.9, 42.8, and 50.3 %, respectively. This demonstrated that hydrophobic interactions took an important role in describing the field properties of GPR35 agonists. Figure 6a illustrates the CoMSIA contour maps of steric and electrostatic fields, which give the similar conclusions as the field distribution of the CoMFA model (Fig. 5).

The hydrophobic contour map of the CoMSIA model in the presence of compound **34** is displayed in Fig. 6b. The white and yellow contour maps highlight areas where

hydrophilic and hydrophobic properties are preferred. A yellow polyhedron was found around the 6-position, which can explain the fact that the bromine substituent is necessary to enhance the biological activity due to the increasing hydrophobicity. There is another large white contour surrounding the benzene ring. This is consistent with the observation that compounds **32–34** with polar substituents in this region display potent activity, whereas compounds **35–36** with more hydrophobic groups (cyclopropylmethoxy and propoxy) exhibit low activity. Moreover, compounds **37** and **38** in which the phenyl ring has been replaced by a pyridine ring have shown better potency compared with compound **29** due to increased hydrophilicity

Topomer CoMFA contour maps

Topomer CoMFA interaction maps (steric and electrostatic interactions) for both fragments (R₁ and R₂) using

compound **34** as a reference structure are shown in Fig. 7. In the steric contour map, green color denotes sterically bulky groups favored for activity, and the yellow color indicates sterically bulky groups unfavored for activity. In the electrostatics contour map, red indicates electronegative favored groups and blue indicates electropositive favored group.

As shown in Fig. 7a, there is a blue region near the 6-position of the R₁ fragment which means sterically bulky group such as bromine is favorable and increase the activity. For the R₂ fragment, green contour (Fig. 7c) surrounding the phenyl ring demonstrated that the introduction of a moderately bulky group into the benzene ring can improve the agonist effects. Meanwhile, two small yellow contours were found outside of the 14-position of R₂ fragment, which can explain well that compounds **35** and **36** containing larger substituents at this position have relatively lower activity.

The electrostatic contour map of the R₁ fragment (Fig. 7b) revealed that electron-withdrawing nature of bromine group substituted at 6-position on the chromone ring was favored for activity. On the other hand, the blue area was found outside of the hydrogen of carboxyl group, which means that this electropositive hydrogen could retain molecular activity. Figure 7d shows two regions of red polyhedral space, indicating that the strong electron-withdrawing group at R₂ resulted in high activity (compounds **26** and **27**). Additionally, the blue polyhedron in the Fig. 7d suggested that electron-rich substituent will reduce the biological activity, which is similar to CoMFA electrostatic contour maps.

Comparison of the CoMFA, CoMSIA, and Topomer CoMFA models

The CoMFA and CoMSIA models were developed based on the alignment of common substructure. Both CoMFA and CoMSIA models have provided the significant correlations of biological activities with structural descriptors. In comparison, the CoMSIA model yielded slightly lower statistical values than the CoMFA model. However, CoMSIA model can give more information because steric, electrostatic, and hydrophobic fields were considered. The CoMFA steric map indicated that a bulkier group around 6 and 14-position will be favorable for higher activity, while the CoMFA electrostatic contour map highlights the importance of the negative substituents at 6-position and positive groups at 14-position for improved potency. The CoMSIA hydrophobic map points out hydrophobic groups at 6-position can enhance the biological activity, whereas the relative hydrophilic substituent around the benzene ring is required for high activity. Topomer CoMFA, the combination of the universal “Topomer” methodology and

CoMFA technologies, belongs to one of the fragment-based 3D-QSAR methods (Ding *et al.*, 2013). In the Topomer CoMFA, fragments can automatically generate based on molecular 3D-pose, and experientially complete identification and alignment of pose of the fragments. The Topomer CoMFA displayed the best statistical parameters in terms of q^2 and r^2 values among these three models. In addition, the contour maps obtained from the Topomer CoMFA model further confirmed the role of steric and electrostatic effects in the agonist activities.

Conclusions

In this study, CoMFA, CoMSIA, and Topomer CoMFA 3D-QSAR models were developed for a series of 8-substituted chromen-4-one-2-carboxylic acid derivatives as GPR35 agonists. All the three models had good statistical results in terms of q^2 and r^2 values, and CoMFA together with Topomer CoMFA models had better predictive ability than CoMSIA model. Furthermore, CoMFA, CoMSIA, and Topomer CoMFA contour maps offered enough information for us to understand 3D-QSAR relationship between structures and their biological activities, which may be considered as a powerful tool in designing and forecasting more efficacious analogs.

Acknowledgments The authors would like to acknowledge the financial supports from the National Natural Science Foundation of China (No. 81101687), the “Innovative Drug Development” State Key Science and Technology of China (No. 2009ZX09103-104), Program for the Top Science and Technology Innovation Teams of Higher Learning Institutions of Shanxi Province, and the Technology Innovation Team of Shanxi Province.

Conflict of interest The author declares that there are no conflicts of interest.

References

- Chandrasekaran V, McGaughey GB, Cavallito CJ, Bowen JP (2004) Three-dimensional quantitative structure–activity relationship (3D-QSAR) analyses of choline acetyltransferase inhibitors. *J Mol Graph Model* 23:69–76
- Cho SJ, Tropsha A (1995) Cross-validated R₂-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. *J Med Chem* 38:1060–1066
- Clark M, Cramer RD, Vanopdenbosch N (1989) Validation of the general-purpose tripos 5.2 force-field. *J Comput Chem* 10:982–1012
- Cosi C, Mannaioni G, Cozzi A, Carla V, Sili M, Cavone L, Maratea D, Moroni F (2011) G-protein coupled receptor 35 (GPR35) activation and inflammatory pain: studies on the antinociceptive effects of kynurenic acid and zaprinast. *Neuropharmacology* 60:1227–1231
- Cramer RD (2012) R-group template CoMFA combines benefits of “ad hoc” and topomer alignments using 3D-QSAR for lead optimization. *J Comput Aided Mol Des* 26:805–819

- Ding W, Sun M, Luo S, Xu T, Cao Y, Yan X, Wang Y (2013) A 3D QSAR study of betulinic acid derivatives as anti-tumor agents using topomer CoMFA: model building studies and experimental verification. *Molecules* 18:10228–10241
- Funke M, Thimm D, Schiedel AC, Müller CE (2013) 8-Benzamido-chromen-4-one-2-carboxylic acids: potent and selective agonists for the orphan G protein-coupled receptor GPR35. *J Med Chem* 56:5182–5197
- Heikamp K, Bajorath J (2013) The future of virtual compound screening. *Chem Biol Drug Des* 81:33–40
- Jenkins L, Alvarez-Curto E, Campbell K, de Munnik S, Canals M, Schlyer S, Milligan G (2011) Agonist activation of the G protein-coupled receptor GPR35 involves transmembrane domain III and is transduced via $G\alpha_{13}$ and β -arrestin-2. *Br J Pharmacol* 162:733–748
- Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 37:4130–4146
- Mackenzie AE, Lappin JE, Taylor DL, Nicklin SA, Milligan G (2011) GPR35 as a novel therapeutic target. *Front Endocrinol (Lausanne)* 2:1–10
- Milligan G (2002) Strategies to identify ligands for orphan G-protein-coupled receptors. *Biochem Soc Trans* 30:789–793
- Min KD, Asakura M, Nakamaru K, Liao Y, Okazaki H, Takahashi T, Fujimoto K, Ito S, Takahashi A, Asanuma H, Yamazaki S, Minamino T, Sanada S, Seguchi O, Nakano A, Ando Y, Otsuka T, Furukawa H, Isomura T, Takashima S, Mochizuki N, Kitakaze M (2010) Identification of genes related to heart failure using global gene expression profiling of human failing myocardium. *Biochem Biophys Res Commun* 393:55–60
- Neetoo-Isseljee Z, Mackenzie AE, Southern C, Jerman J, McIver EG, Harries N, Taylor DL, Milligan G (2013) High-throughput identification and characterization of novel, species-selective GPR35 agonists. *J Pharmacol Exp Ther* 344:568–578
- Ohshiro H, Tonai-Kachi H, Ichikawa K (2008) GPR35 is a functional receptor in rat dorsal root ganglion neurons. *Biochem Biophys Res Commun* 365:344–348
- Pal K, Melcher K, Xu HE (2012) Structure and mechanism for recognition of peptide hormones by Class B G-protein-coupled receptors. *Acta Pharmacol Sin* 33:300–311
- Taniguchi Y, Tonai-Kachi H, Shinjo K (2006) Zaprinast, a well-known cyclic guanosine monophosphate-specific phosphodiesterase inhibitor, is an agonist for GPR35. *FEBS Lett* 580:5003–5008
- Thimm D, Funke M, Meyer A, Müller CE (2013) 6-Bromo-8-(4-[(3H)methoxybenzamido]-4-oxo-4H-chromene-2-carboxylic acid: a powerful tool for studying orphan G protein-coupled receptor GPR35. *J Med Chem* 56:7084–7099
- Wang J, Simonavicius N, Wu X, Swaminath G, Reagan J, Tian H, Ling L (2006) Kynurenic acid as a ligand for orphan G protein-coupled receptor GPR35. *J Biol Chem* 281:22021–22028
- Wang H, Du Z, Zhang C, Tang Z, He Y, Zhang Q, Zhao J, Zheng X (2014) Biological evaluation and 3D-QSAR studies of curcumin analogues as aldehyde dehydrogenase 1 inhibitors. *Int J Mol Sci* 15:8795–8807
- Wilson GL, Lill MA (2011) Integrating structure-based and ligand-based approaches for computational drug design. *Future Med Chem* 3:735–750
- Yang Y, Lu JY, Wu X, Summer S, Whoriskey J, Saris C, Reagan JD (2010) G-protein-coupled receptor 35 is a target of the asthma drugs cromolyn disodium and nedocromil sodium. *Pharmacology* 86:1–5